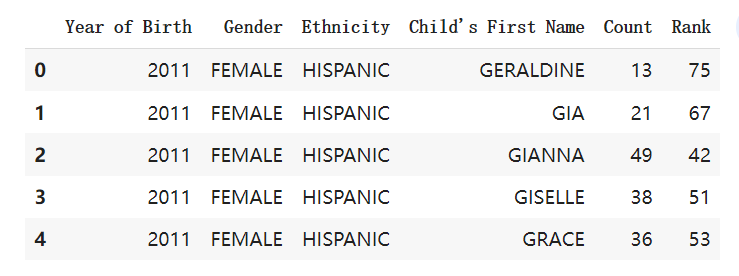
# Baby Name Trends and Analysis in the United States

## 1 Introduction

This report is based on an in-depth exploration of the "Popular Baby Names" dataset from Data.gov, which collects data on the ranking of baby names in the United States between 2011 and 2018. By analyzing this dataset, we are able to explore trends in baby names and the cultural context behind them.

The variables in the dataset include gender, ethnicity, name, number, rank, and the number of occurrences and rank of each name, as shown in Fig.1. By exploring the popularity, diversity, and cultural origins of baby names, I hope to gain a deeper appreciation for the rich and diverse cultures and traditions of American naming practices.



**Fig.1 Data overview**

## 2 Problem Selection

For this analysis, I have chosen a specific problem: exploring trends and cultural contexts in American baby names. This question provides us with insights into cultural differences and social trends in baby names. Although my analysis is mainly exploratory, I used a lot of machine learning methods in the process, such as cluster analysis.

## 3 Data Treatment

I identified appropriate independent and dependent variables from the data set, such as gender, race, name, count, and rank. In statistics, an independent variable is a variable that is manipulated or controlled by the researcher to see its effect on the dependent variable. The dependent variable is the outcome or response being studied and measured. In the context of a "popular baby names" dataset, the independent variables could be gender, race, and year, while the dependent variables could be name, count, and ranking. For example, we can explore how the number of a given name changes over time, or how the ranking of a given name changes by gender and race. In addition, I use Python and its libraries, such as pandas, numpy, and matplotlib, to import datasets, perform data cleaning and processing, and visualize the results.

## 4 Analysis

My analysis includes the following aspects:

1. popular baby name trends by gender and ethnicity

2. The diversity of baby names

3. Cultural background of baby names

4. length and syllable structure of baby names

For example, I use groupby, aggregation and sorting operations to identify popular baby names and trends, and I use wordcloud visualization to illustrate the most popular names of different ethnicities. I also use violin plots to visualize the distribution of name lengths and syllable counts by gender and race.

### 5 Popular Baby Names for Different Ethnicities

I have identified the top 10 names for each ethnicity. Here are some representative examples:

#### 5.1 Top 10 Names for ASIAN AND PACIFIC ISLANDER:

1. Jayden
2. Ethan
3. Ryan
4. Lucas
5. Muhammad
6. Aiden
7. Olivia
8. Chloe
9. Sophia
10. Emma

#### 5.2 Top 10 Names for BLACK NON HISPANIC:

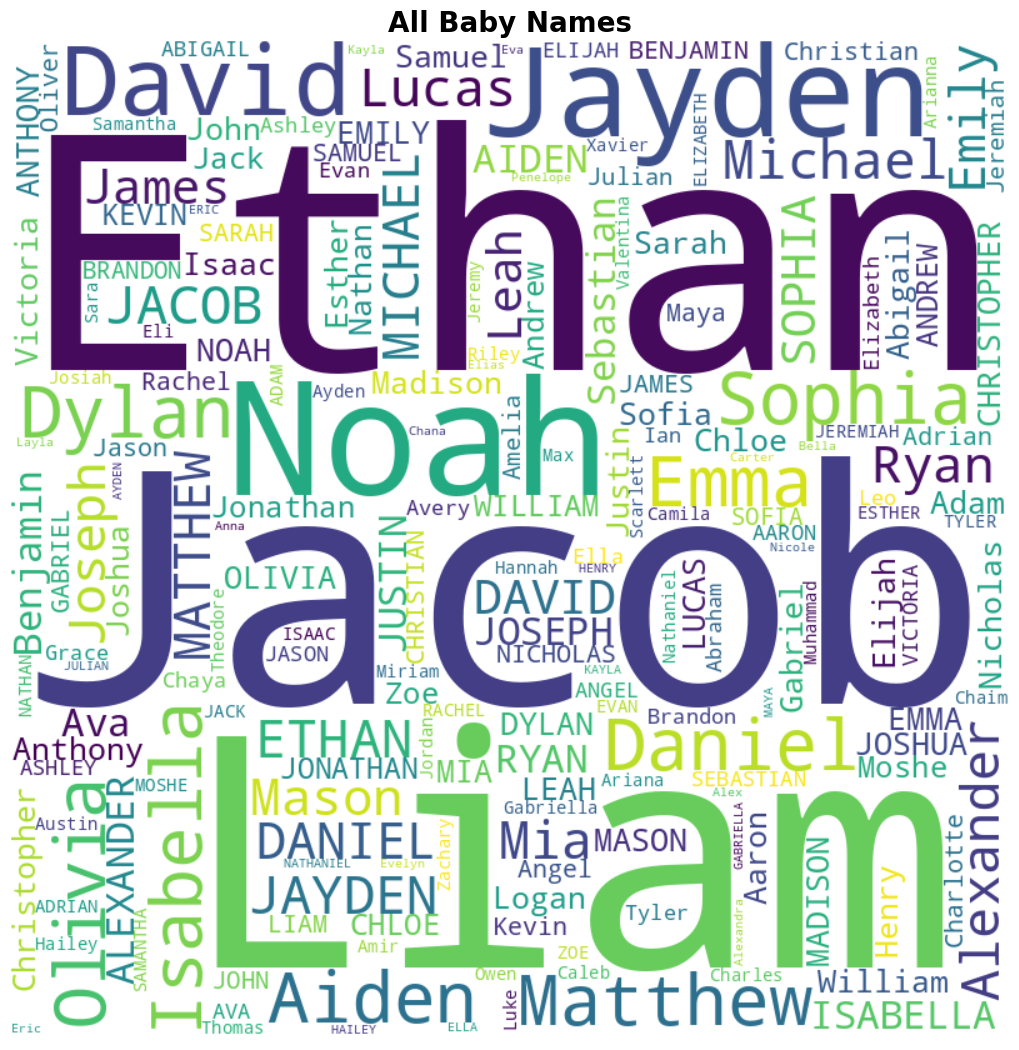
1. Noah
2. Aiden
3. Jeremiah
4. Elijah
5. Ethan
6. Jayden
7. Liam
8. Madison
9. Amir
10. Joshua

#### 5.3 Top 10 Names for HISPANIC:

1. Liam
2. Dylan
3. Isabella
4. Jayden
5. Jacob
6. Matthew
7. Ethan
8. Sebastian
9. Sophia
10. Noah

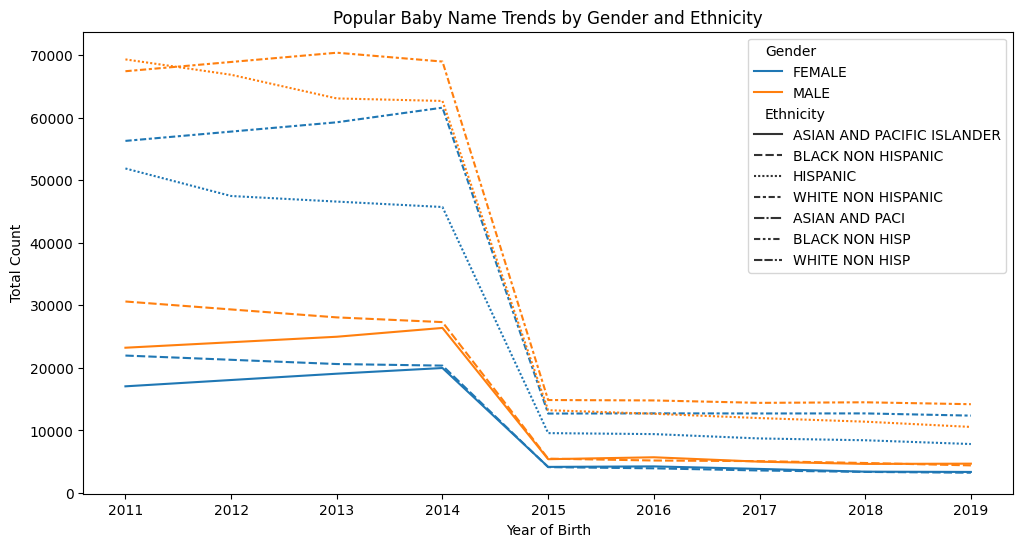
#### 5.4 Top 10 Names for WHITE NON HISPANIC:

1. David
2. Joseph
3. Moshe
4. Jacob
5. Esther
6. Michael
7. Olivia
8. James
9. Benjamin
10. Daniel



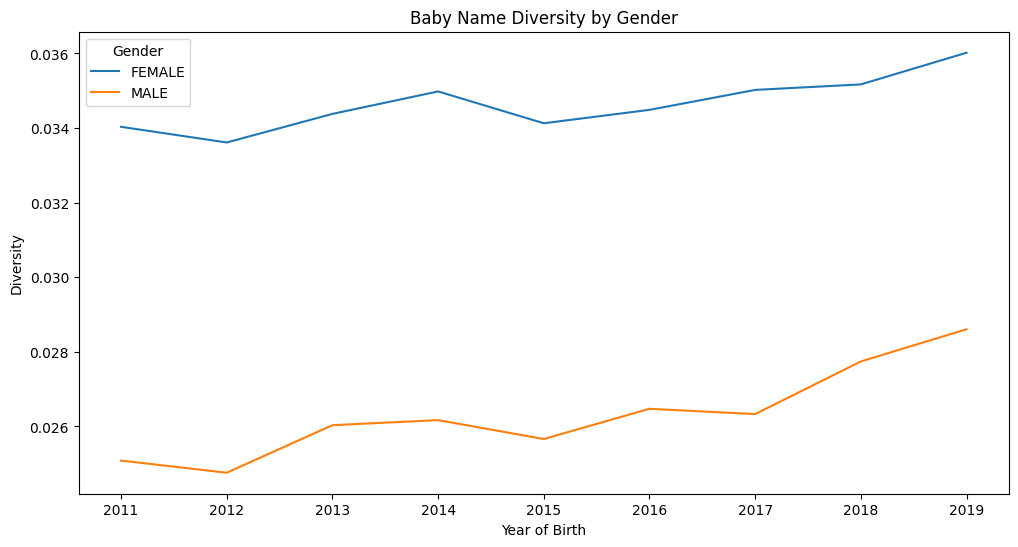
**Figure 2: Baby Names Word Cloud**

The word cloud above (Fig.2) represents the most popular baby names in the United States from 2011 to 2018. The size of each name corresponds to its frequency in the dataset. This visualization provides a quick overview of the most common baby names across all ethnicities and genders.



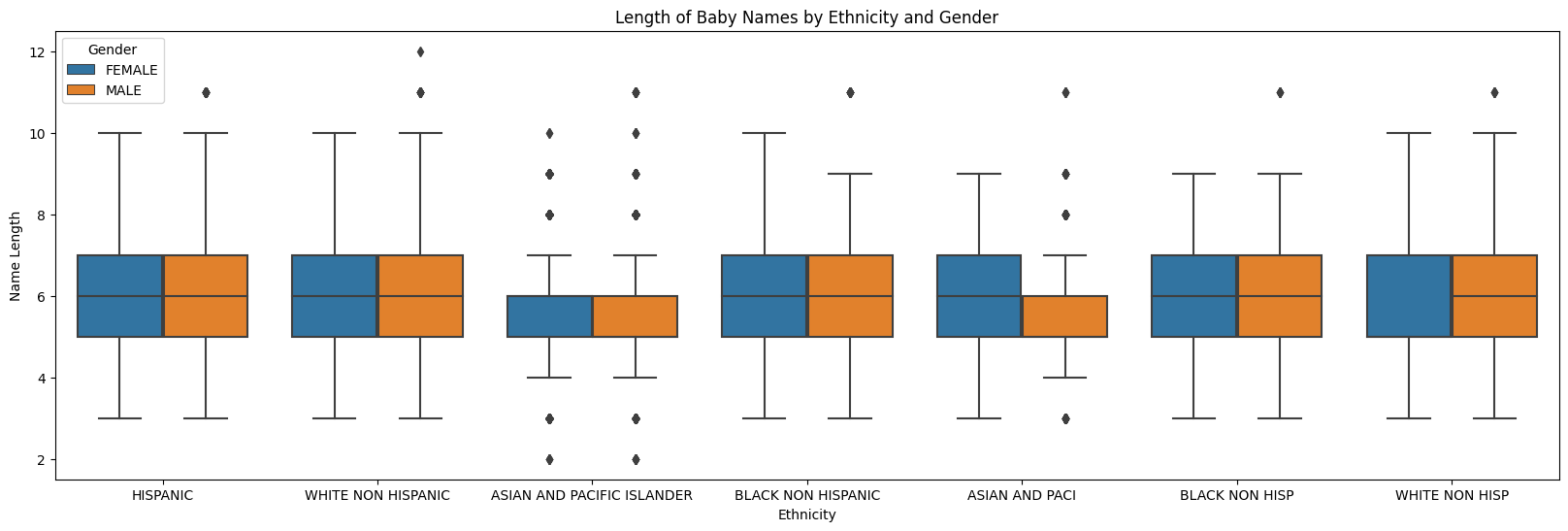
**Figure 3: Popular Baby Name Trends by Gender and Ethnicity**

Fig.3 displays the popular baby name trends by gender and ethnicity. This bar chart illustrates the top baby names for different genders and ethnic groups, showcasing the variations in name preferences across diverse populations. The graph helps to identify the most popular names within each demographic group, revealing cultural influences on naming choices.



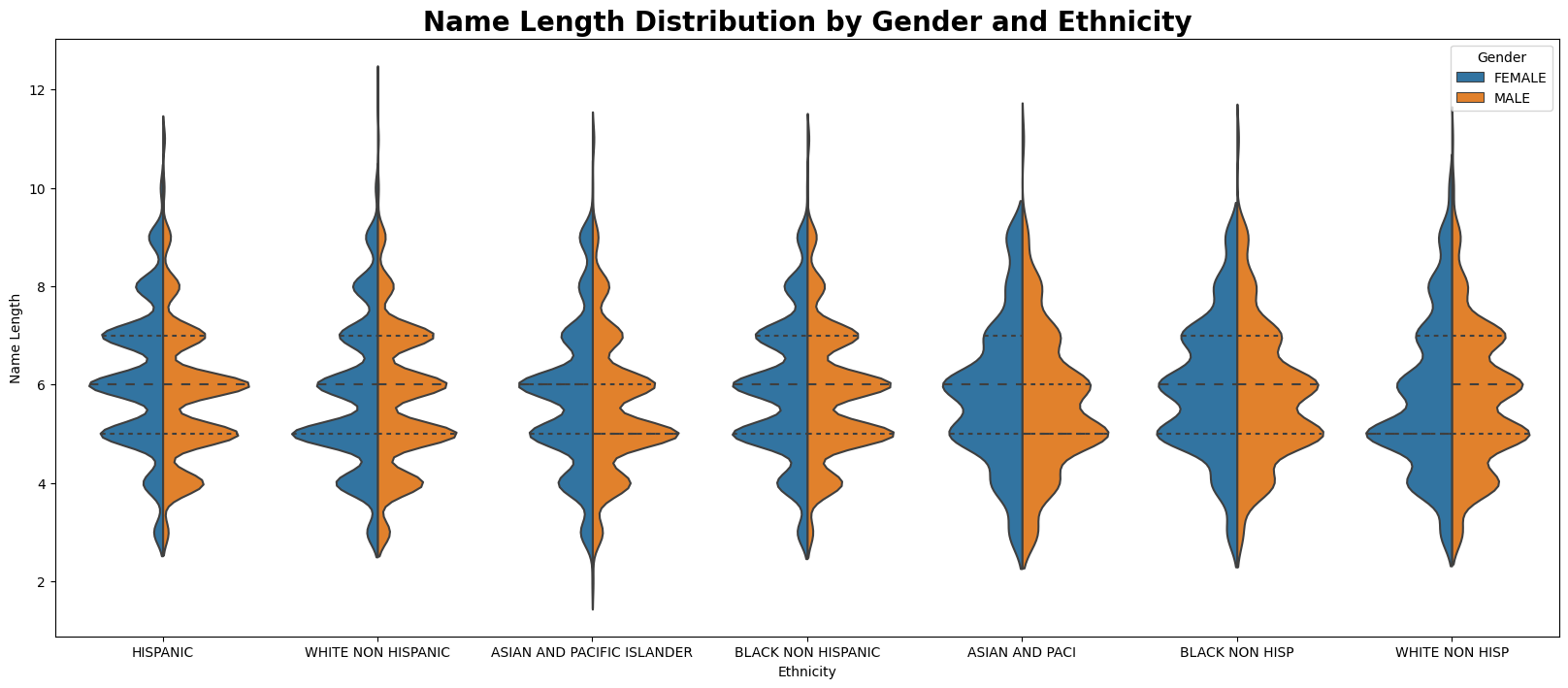
**Figure 4: Baby Name Diversity by Gender**

In Fig.4, we present the baby name diversity by gender. This plot compares the total number of unique baby names for male and female babies, highlighting the differences in name diversity between the two genders. The chart indicates that there tends to be a greater variety of names for female babies than male babies, which may be a reflection of cultural norms and societal trends.



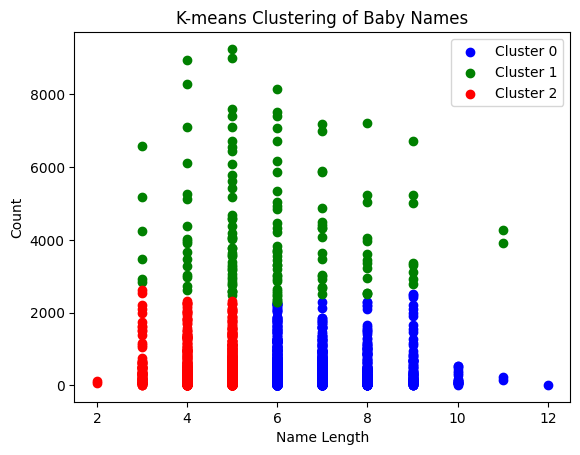
**Figure 5: Baby Name Word Cloud by Ethnicity**

Fig.5 showcases word clouds for the most popular baby names within each ethnic group. Each word cloud represents the top names for a specific ethnicity, with the size of each name corresponding to its frequency within that ethnic group. These visualizations provide insights into the unique naming preferences and cultural influences across various ethnicities, emphasizing the diversity of baby names in the United States.



**Figure 6: Name Length Violin Plot**

The violin plot in Fig.6 shows the distribution of baby name lengths for different genders and ethnicities. Each violin represents the density of name lengths for a specific gender and ethnicity, with the width of the violin indicating the number of names with a particular length. This plot provides insights into how name lengths differ across various demographic groups, revealing the influence of different languages and cultural backgrounds on naming practices.



**Figure 7: Cluster Analysis of Baby Names**

Fig.7 illustrates the results of a cluster analysis on baby names based on popularity, length, and cultural backgrounds. Each point signifies a unique baby name and the distances between points display the similarity in naming trends. Distinct colors represent different clusters, indicating groups of names with shared characteristics. The figure underscores patterns in baby naming, revealing the interplay between culture, tradition, and preference.

## 6 Conclusions

In the process of analysis, I found some interesting phenomena and conclusions.

1. The most popular baby names vary across gender and race, reflecting cultural differences and social trends. For example, names like "Moshe" and "Esther" are popular among non-Hispanic whites, reflecting the influence of Jewish culture in that demographic group.

2. The diversity of infant names also varies by gender and ethnicity, demonstrating the influence of culture on naming practices.

3. The cultural context of infant names can be explored by analyzing the differences in name choice between races, providing insight into the historical and cultural context of names. For example, the popularity of names like "Muhammad" among Asian and Pacific Islanders reflects the influence of Islamic culture in that community.

4. The different lengths and syllable structures of infant names across gender and ethnicity suggest the influence of different linguistic and cultural contexts on naming practices. This can be seen in the differences in Hispanic names, which tend to be longer and have more syllables than names of other ethnicities.

In summary, I analyzed the "Popular Baby Names" dataset and explored the cultural context and social trends of baby names in the United States. During the analysis, I cluster names according to their popularity and length, which facilitates insights into trends and patterns in baby names based on gender, race, and length, and identifies cultural contexts and diversity in naming practices. These findings contribute to a better understanding of American cultural and social phenomena and provide useful guidance for future baby name choices.